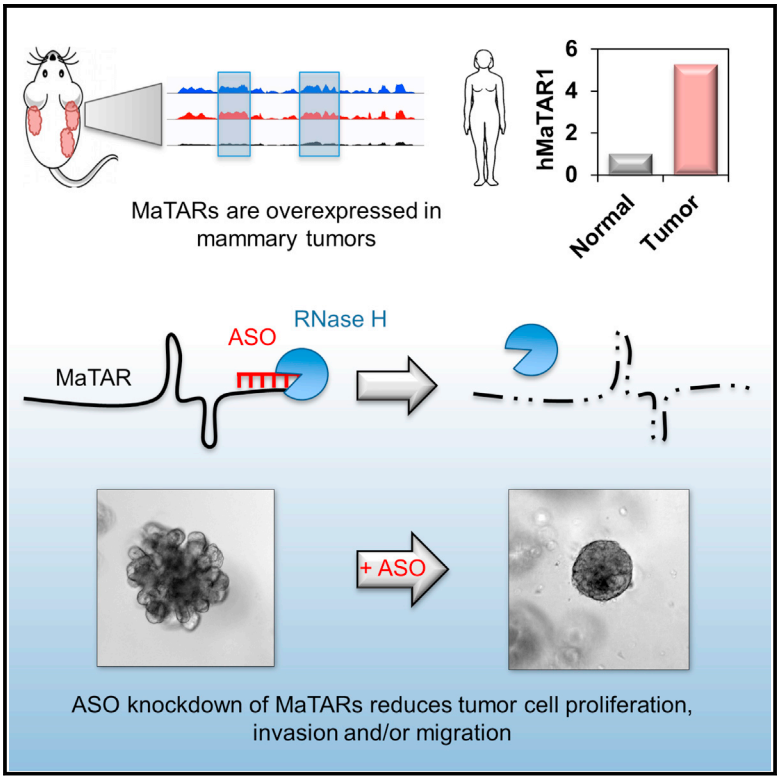
Article

Cell Reports

Mammary Tumor-Associated RNAs Impact Tumor Cell Proliferation, Invasion, and Migration

Graphical Abstract



Authors

Sarah D. Diermeier, Kung-Chi Chang, Susan M. Freier, ..., Frank Rigo, C. Frank Bennett, David L. Spector

Correspondence

spector@cshl.edu

In Brief

Diermeier et al. identify differentially expressed lncRNAs in mouse models of breast cancer. They prioritize potentially oncogenic lncRNAs (mammary-tumor-associated RNAs or MaTARs), validate these using antisense knockdown and identify their human counterparts (*hMaTARs*). A number of MaTARs impact cell proliferation, invasion, and/or migration.

Highlights

- Hundreds of lncRNAs are dysregulated in mouse mammary tumors
- 30 overexpressed lncRNAs were prioritized as mammary-tumor-associated RNAs (MaTARs)
- 20 MaTARs impact cancer cell proliferation, invasion, and/or organoid branching
- *hMaTARs* are overexpressed in human breast tumors and might be of clinical relevance

Accession Numbers

GSE72823

Diermeier et al., 2016, Cell Reports *17*, 261–274
September 27, 2016
http://dx.doi.org/10.1016/j.celrep.2016.08.081

Cell Reports
Article

Mammary Tumor-Associated RNAs Impact Tumor Cell Proliferation, Invasion, and Migration

Sarah D. Diermeier,[1] Kung-Chi Chang,[1,2] Susan M. Freier,[3] Junyan Song,[1,4] Osama El Demerdash,[1] Alexander Krasnitz,[1,4] Frank Rigo,[3] C. Frank Bennett,[3] and David L. Spector[1,2,5,*]
[1]Cold Spring Harbor Laboratory, Cold Spring Harbor, NY 11724, USA
[2]Molecular and Cellular Biology Program, Stony Brook University, Stony Brook, NY 11794, USA
[3]Ionis Pharmaceuticals, Carlsbad, CA 92010, USA
[4]Department of Applied Mathematics and Statistics, Stony Brook University, Stony Brook, NY 11794, USA
[5]Lead Contact
*Correspondence: spector@cshl.edu
http://dx.doi.org/10.1016/j.celrep.2016.08.081

SUMMARY

Long non-coding RNAs (lncRNAs) represent the largest and most diverse class of non-coding RNAs, comprising almost 16,000 currently annotated transcripts in human and 10,000 in mouse. Here, we investigated the role of lncRNAs in mammary tumors by performing RNA-seq on tumor sections and organoids derived from MMTV-PyMT and MMTV-Neu-NDL mice. We identified several hundred lncRNAs that were overexpressed compared to normal mammary epithelium. Among these potentially oncogenic lncRNAs we prioritized a subset as Mammary Tumor Associated RNAs (MaTARs) and determined their human counterparts, *hMaTARs*. To functionally validate the role of MaTARs, we performed antisense knockdown and observed reduced cell proliferation, invasion, and/or organoid branching in a cancer-specific context. Assessing the expression of *hMaTARs* in human breast tumors revealed that 19 *hMaTARs* are significantly upregulated and many of these correlate with breast cancer subtype and/or hormone receptor status, indicating potential clinical relevance.

INTRODUCTION

A number of studies over the past decade revealed that only 2% of the human genome encodes for proteins, while as much as 80% can be transcribed in a cell-type-specific manner (Bertone et al., 2004; Carninci et al., 2005; Djebali et al., 2012; ENCODE Project Consortium, 2012; Kapranov et al., 2007; Katayama et al., 2005; Okazaki et al., 2002). Non-coding transcripts are subdivided into several classes, with long non-coding RNAs (lncRNAs) representing the largest and most diverse class. LncRNAs are defined by length, ranging from 200 nt to 138 kb. Like mRNAs, they are transcribed by RNA polymerase II and can be capped, spliced, and poly-adenylated. However, lncRNAs lack a significant open reading frame, can be transcribed from the sense or antisense orientation and commonly originate from introns or intergenic regions (Derrien et al., 2012; Iyer et al., 2015; St Laurent et al., 2012). Members of this class of non-coding transcripts have been implicated as regulatory molecules in a variety of cellular functions (for review, see Bergmann and Spector, 2014; Kornienko et al., 2013; Rinn and Chang, 2012).

In recent years, lncRNAs emerged as a novel class of regulatory molecules in cancer (for review, see Cheetham et al., 2013; Prensner and Chinnaiyan, 2011; Shore et al., 2012; Wahlestedt, 2013; Wapinski and Chang, 2011). One of the first examples of a lncRNA that has been studied in the context of breast cancer is the HOX antisense intergenic RNA (HOTAIR) (Rinn et al., 2007). It promotes mammary tumor invasion as well as metastasis and acts as an independent predictor of patient survival rates (Gupta et al., 2010; Rinn et al., 2007). Elevated expression levels of another lncRNA BCAR4 were described to correlate with a higher rate of metastasis and shorter patient survival (Xing et al., 2014). Recently, lncRNA152 and lncRNA67 were shown to play a role in the growth of breast cancer cell lines (Sun et al., 2015). Eleanors, a new class of lncRNAs, were found to be associated with breast cancer adaptation (Tomita et al., 2015). In addition, genetic knockout or ASO-mediated knockdown of *Malat1* results in the differentiation of mammary tumors and significant reduction of metastasis in vivo (Arun et al., 2016). Differential expression of a few additional lncRNAs has been studied in the context of breast cancer, but as for most lncRNAs the functional mechanisms remain elusive (for review, see Hansji et al., 2014; Vikram et al., 2014). As only a handful of the almost 16,000 annotated lncRNAs (GENCODE v.24) have been functionally characterized and studied in the context of breast cancer, it is essential to perform an unbiased RNA sequencing (RNA-seq) screen to identify the complement of lncRNAs that exhibit altered expression in mammary tumors.

Here, we generated a comprehensive compendium of lncRNAs that are differentially expressed in primary mammary tumors compared to the normal mammary gland epithelium. We performed RNA-seq analysis on three physiologically relevant transgenic mouse models representing the luminal B (MMTV-PyMT) and the HER2/neu-amplified (MMTV-Neu-NDL and MMTV-Cre;Flox-Neo-Neu-NT) subtypes of breast cancer (Andrechek et al., 2000; Guy et al., 1992; Siegel et al., 1999). Among the 290 lncRNAs that are upregulated at least 2-fold in

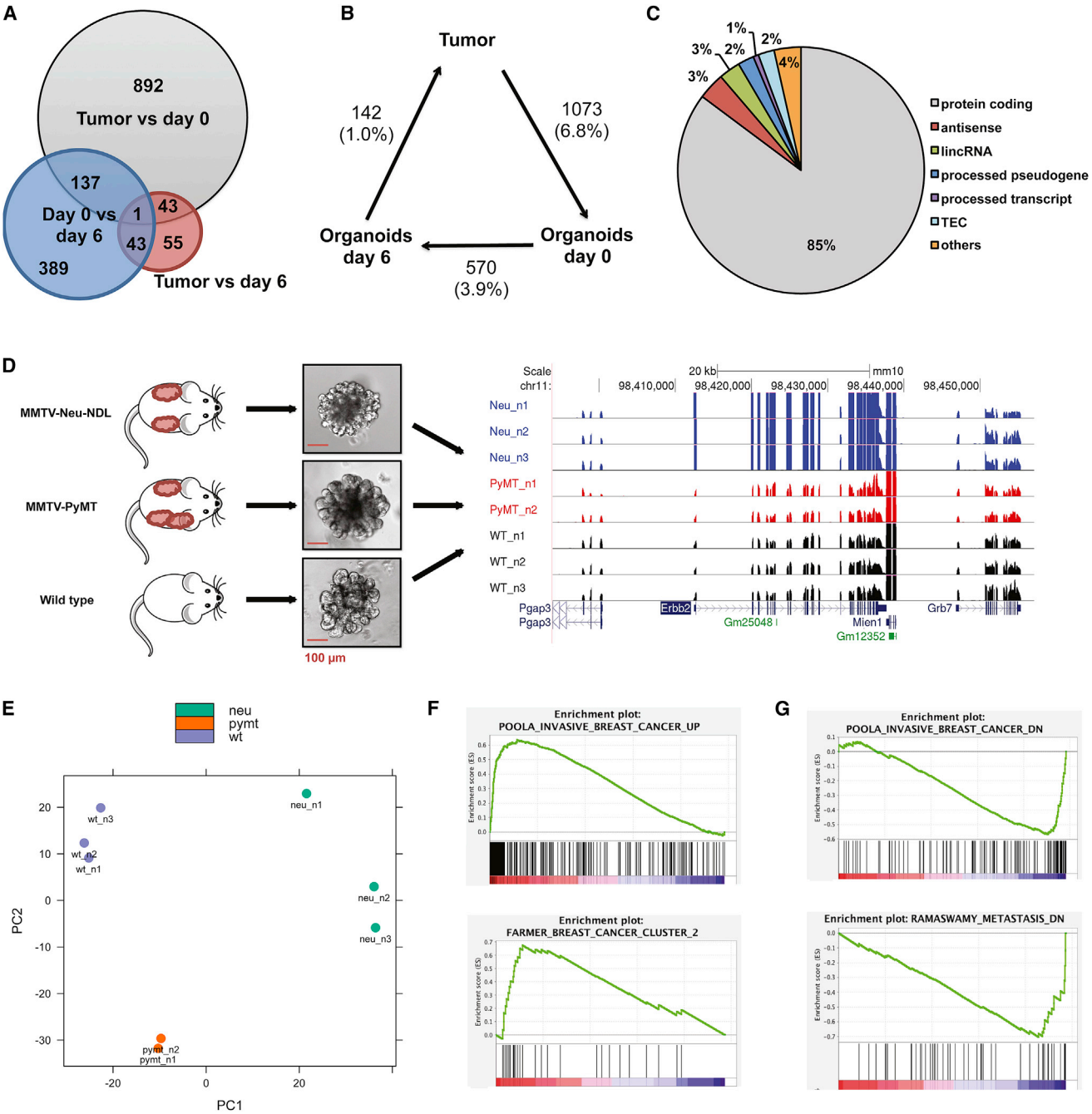


Figure 1. The Transcriptome of MMTV-PyMT and MMTV-Neu-NDL Tumors

(A) Venn diagram indicating the overlap of differentially expressed genes across all datasets.

(B) Illustration of the number of genes significantly altered comparing tumor, day 0 and day 6 organoid datasets. Percentage of the transcriptome is denoted in brackets; i.e., 142 genes represent 1.0% of the transcriptome in day 6 organoids compared to tumor sections.

(C) Transcript biotypes of the 142 differentially expressed genes comparing day 6 organoids to tumors. The majority (85%) of altered transcripts are protein-coding. TEC, to be experimentally confirmed (GENCODE classification).

(D) Organoids were prepared from MMTV-PyMT, MMTV-Neu-NDL, and WT normal mammary epithelia and cultured for 6 days in Matrigel (left). UCSC genome browser image (right) displays RNA-seq reads exemplarily at the *Erbb2/neu* locus. Vertical viewing range: 0–50.

(E) Principal component analysis comparing the transcriptome of organoids derived from WT (purple), MMTV-PyMT tumors (orange), and MMTV-Neu-NDL tumors (green). Three clusters were detected according to the three different mouse models.

(legend continued on next page)

the tumors compared to normal mammary epithelial cells we prioritized 30 previously uncharacterized lncRNAs as mammary-tumor-associated RNAs (MaTARs). In order to functionally validate the MaTARs, we performed antisense knockdown studies in primary cancer cells and 3D ex vivo mammary tumor organoids. Our results indicate that independent knockdown of 20 of the 30 investigated MaTARs significantly reduce mammary tumor cell proliferation, invasion, and/or collective cell migration. Notably, we observed phenotypic changes upon MaTAR knockdown in organoids derived from tumors but not in organoids derived from normal mammary epithelial cells, indicating the cancer-specific function of MaTARs. Furthermore, we confirmed the overexpression of 19 human MaTARs (*hMaTARs*) comparing breast tumors to matched normal controls. Of those, 11 are expressed in a subtype-specific manner. We propose that these previously uncharacterized lncRNAs might act as prognostic indicators and/or drivers of breast tumor progression.

RESULTS

The Transcriptome of Luminal B and HER2-Amplified Mammary Tumors

Recently, the culturing of mammary organoids in 3D artificial extracellular matrix (ECM) hydrogels has gained popularity over 2D cell culture approaches, especially for studying mammalian development and disease (for review, see Clevers, 2016; Shamir and Ewald, 2014). Previous studies demonstrated that mammary branching morphogenesis can be recapitulated in an organoid system by retaining its epithelial spatial organization (Barcellos-Hoff et al., 1989; Ewald et al., 2008; Fata et al., 2007; Nguyen-Ngoc et al., 2012). Furthermore, organoids developed from tumors have the potential to serve as a rapid in vitro screening system in drug development and "personalized medicine."

We set out to validate that organoids indeed closely resemble the gene expression signature of the tumors they were derived from. Therefore, we removed mammary tumors from three MMTV-PyMT mice and used one-half of each tumor for direct RNA isolation from the tissue sections. The other half was used to generate organoids, and RNA was extracted either immediately following organoid preparation (day 0) or after 6 days of culturing in 3D Matrigel domes in the presence of FGF2 (day 6). Comparative RNA-seq analyses revealed that the global gene expression signature of mammary tumor organoids is comparable to the tumor transcriptome (Figures S1A, S1B, and S2A). Importantly, only 1% of the transcriptome (142 out of the 13,854 expressed genes) showed statistically significant alterations (Figures 1A, 1B, and S1C) between the tumor and the day 6 organoids. The affected genes were almost exclusively (85%) protein coding genes (Figures 1C and S1D). Several pathways involved in signaling and metabolism appeared altered when taking into account genome-wide fluctuations (Figure S2B).

As our initial comparison demonstrated that organoids are a valid model system to study lncRNA expression, we generated organoids from MMTV-PyMT and MMTV-Neu-NDL mice. After 6 days in Matrigel, the transcriptome of the organoids was analyzed using RNA-seq. To classify cancer-specific changes in the gene expression signature in comparison to normal tissue, we also generated and sequenced organoids from nulliparous mammary glands of three wild-type friend virus B (FVB) females (Figure 1D). The RNA-seq screen revealed significant changes in both the coding and the long non-coding transcriptome, with a total of 4,633 genes differentially expressed in MMTV-PyMT- and 4,322 in MMTV-Neu-NDL-derived organoids. As an internal control, we monitored the expression of *Erbb2/neu*, which was overexpressed only in the MMTV-Neu-NDL organoids but not in the PyMT or wild-type (WT) organoids (Figures 1D and S1F). Principal component analysis and Euclidean distance plots comparing WT mammary glands to the luminal B and HER2/neu-dependent mammary tumors revealed distinct clustering according to the tumor-driving transgene (Figures 1E and S1E). Furthermore, we observed general and subtype-specific expression changes in cellular pathways (Figure S2C). Importantly, gene-set enrichment analyses (GSEA) revealed that our mouse mammary tumor organoid data resembled the transcriptome of human breast cancer. For instance, the downregulated genes of both models correlated well with expression signatures of invasive breast cancer and metastatic tumors. A similar correlation was found for upregulated genes in both datasets (Figures 1F and 1G).

Taken together, these data demonstrate that mouse mammary carcinoma-derived organoids recapitulate the transcriptome of luminal B and HER2/neu-amplified breast cancer and are a reliable model system to study the expression of non-coding RNAs.

Identification of Upregulated lncRNAs in Mammary Organoids

We set out to generate a compendium of all currently annotated lncRNAs that exhibit altered expression levels in mammary tumors to provide a useful database for future research. We identified 484 lncRNAs in MMTV-PyMT and 402 lncRNAs in MMTV-Neu-NDL with an overlap of 122 between both models that are differentially expressed in organoids derived from mammary tumors compared to WT mammary epithelial cells (Figure 2A). These genes represent a diverse group in terms of their biotypes and genomic location (Figures S3A and S3B). Biotypes include "lincRNA" (long intergenic non-coding RNAs), "processed pseudogenes," "antisense" transcripts, and "processed transcripts" (Figure S3A).

Furthermore, we aimed to focus on potentially oncogenic lncRNAs that might act as drivers of tumor progression and filtered our dataset for lncRNAs that are upregulated at least 2-fold. We identified 109 upregulated lncRNAs in MMTV-PyMT (Table S1) and 207 in MMTV-Neu-NDL organoids (Table S2),

(F) Examples of GSEA datasets found upregulated in MMTV-Neu-NDL tumors (upper panel) and MMTV-PyMT tumors (lower panel). FDR < 0.01.
(G) Examples of GSEA datasets found downregulated in MMTV-Neu-NDL tumors (upper panel) and both MMTV-Neu-NDL and MMTV-PyMT tumors (lower panel). FDR < 0.01.
See also Figures S1 and S2 and Tables S1 and S2.

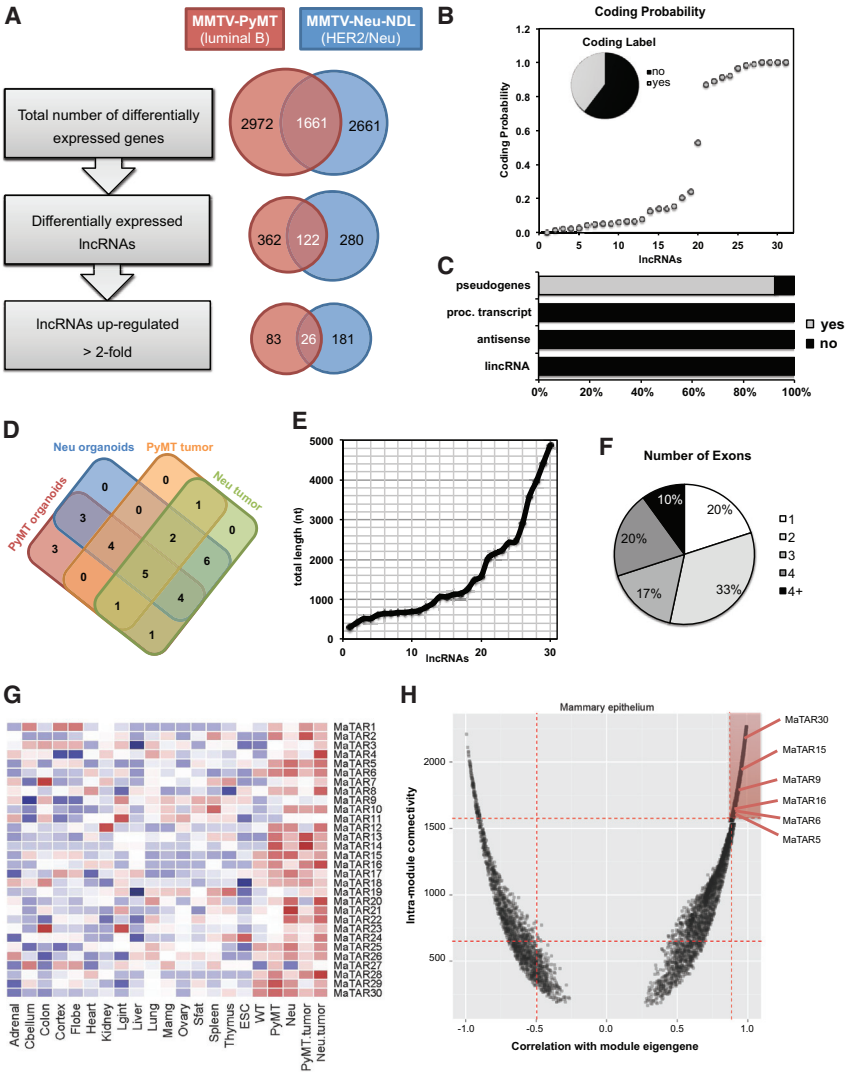


Figure 2. Identification and Characterization of MaTARs

(A) Computational analysis pipeline to identify potentially oncogenic lncRNAs. Numbers indicate genes significantly differentially expressed (FDR adjusted p value < 0.1). Ensembl IDs and log2(fold change) of all upregulated (> 2-fold) genes is provided in Tables S1 and S2.

(B) Coding probability of all 30 MaTARs according to CPAT. Coding probability is plotted on a scale from 0 to 1, with 0 indicating "no coding potential" and "1" representing "high coding potential." Inset: 63% of all MaTARs are non-coding (no), whereas 37% are potentially protein-coding (yes).

(C) Correlation of the CPAT coding label and lncRNA biotypes. The 37% of transcripts with protein-coding potential in B) are attributed to the pseudogene fraction of MaTARs.

(D) Venn diagram illustrating overexpression of the 30 MaTARs in four tumor models. PyMT tumor, tumor sections from MMTV-PyMT sections; Neu tumor, tumor sections from MMTV-Cre;Flox-Neo-Neu-NT. Most MaTARs are upregulated in several datasets.

(E) Length distribution of all 30 MaTARs. The majority (20 MaTARs) is < 1,500 nt long.

(F) Number of exons per MaTAR. More than 50% of MaTARs consist of one or two exons; only three transcripts comprise more than four exons.

(G) MaTAR expression (variance-stabilizing transformation [vst] counts) across ENCODE tissues, wild-type (WT) organoids, and tumor organoids and sections. The highest expression level for 18 MaTARs was detected in tumors. All MaTARs are expressed in a tissue-specific manner. PyMT. tumor, tumor sections from MMTV-PyMT sections; Neu.tumor, tumor sections from MMTV-Cre;Flox-Neo-Neu-NT. Counts are scaled per row.

(H) Connectivity between genes in the "mammary epithelium" module is plotted against the module's eigengene correlation. Red dotted lines correspond to first and third quartiles. Six MaTARs are in the fourth quartile, denoted by high eigengene correlation and intra-module connectivity.

See also Figures S1, S3, and S4 and Tables S3 and S4.

with an overlap of 26 lncRNAs (Figure 2A, highlighted in gray in Tables S1 and S2). Interestingly, very few of the identified transcripts have been studied previously. One of these lncRNAs is Trp53cor1 (also known as lincRNA-p21), which is upregulated at least 4-fold in both mammary tumor models (Tables S1 and S2) and has been implicated in p53/p21-dependent gene regulation (Dimitrova et al., 2014; Huarte et al., 2010) as well as cancer (Chou et al., 2015; Yang et al., 2014). Furthermore, the two pluripotency-associated transcripts Platr4 and Platr7 were upregulated more than 4-fold in MMTV-Neu-NDL organoids (Table S2). Notably, Platr lncRNAs were identified in an RNA-seq screen of embryonic stem cells (ESCs) and shown to be correlated to the maintenance of the ESC expression profile (Bergmann et al., 2015). We performed coding potential analysis on all 290 upregulated lncRNA genes (443 associated transcripts in total) using the Coding-Potential Assessment Tool (CPAT) (Wang et al., 2013). Our results revealed that 75% of the 443 identified transcripts do not exhibit coding potential (Figures S3C–S3F) despite including pseudogenes in our analysis. As expected, we observe a good correlation between coding probability and open reading frame (ORF) length but not RNA length (Figures S3E and S3F).

We further characterized the identified differentially expressed lncRNA genes regarding shared transcription factor binding sites at their promoter regions. To do so, we performed motif analyses in the regions of −900 to +100 bp around the transcription start site (TSS) of all differentially regulated lncRNA genes, identified in both PyMT and Her2/neu organoids. For downregulated promoters, we observe the enrichment of a rather rare motif matching Zbtb33/Kaiso as well as motifs for NF-κB and CEBP (Figure S4A). In contrast, upregulated lncRNA genes seem to be primarily regulated by androgen receptor (AR) and p63 (Figure S4B).

To narrow down the long list of candidates to the ones with highest potential to impact mammary tumors, we prioritized the transcripts further based on several stringent criteria: (1) statistical significance (false discovery rate [FDR] adjusted p value

<0.1), (2) at least 2-fold upregulation in tumor organoids and/or tumor sections compared to normal mammary organoids, (3) sufficient read coverage per transcript to eliminate very lowly abundant RNAs (Figure S1G), (4) human conservation based on sequence and/or synteny, (5) location in an intergenic genomic region, and (6) lack of highly repetitive elements. The last two points ensure knockdown specificity, i.e., that antisense oligonucleotides (ASOs) can be designed for the lncRNA of interest while avoiding potential off-target effects. Out of 290 upregulated lncRNAs, 30 transcripts fulfilled the above-mentioned criteria and were selected as an initial subset for further characterization. All 30 are overexpressed in at least four out of the five sequenced PyMT organoid datasets as well as significantly upregulated in PyMT tumor sections (Figure 2D). We also included RNA-seq data from MMTV-Cre;Flox-Neo-Neu-NT tumors, a second model for the HER2/neu-amplified subtype of human breast cancer, to further refine our candidate lncRNAs. We termed the prioritized transcripts Mammary Tumor Associated RNA 1-30 (MaTAR1-30). A complete list of their official gene IDs and expression levels in all evaluated mammary tumor models is presented in Table S3. Of the 30 MaTARs, only MaTAR4 and MaTAR8 have been characterized in the past. MaTAR4 (Hoxa transcript antisense RNA, myeloid-specific 1; Hotairm1) has been implicated in myelopoiesis and myeloid maturation (Zhang et al., 2009; 2014). MaTAR8 (lncRNA-Smad7) has been described as a TGF-β-regulated antisense transcript of *Smad7* inhibiting apoptosis in mouse breast cancer cells (Arase et al., 2014). We tested the coding potential of all MaTARs using CPAT (Wang et al., 2013). While 60% of the MaTARs were classified as non-coding, 40% were predicted to be coding, attributed to the pseudogene fraction (Figures 2B and 2C). We further characterized the MaTARs according to their total length and number of exons (Figures 2E and 2F). MaTARs range in size from 291 to 4867 nt with the majority of MaTARs being shorter than 1,500 nt. More than half of all MaTARs consist of one or two exons, only 10% show complex structures with more than four exons.

In order to compare the expression levels of MaTARs in mammary tumors to other tissue types, we analyzed the ENCODE expression datasets of adult mouse tissue as well as embryonic stem cells (Bergmann et al., 2015) (Figure 2G). Our results confirm the general notion of tissue-specific expression of lncRNAs (Cabili et al., 2011; Derrien et al., 2012), with the majority of the MaTARs (18 out of 30) being expressed at very low levels in tissues other than mammary tumors. The remaining 12 MaTARs (MaTAR1, 3, 7, 9, 10, 11, 12, 18, 19, 23, 26, and 27) are also overexpressed in mammary tumors compared to normal mammary glands, but in addition these MaTARs are present in comparable levels in at least one other tissue type.

To further characterize MaTARs in the context of global gene expression and to assess their potential role in mammary tumor progression, we performed weighted gene correlation network analysis (WGCNA) (Langfelder and Horvath, 2008). We identified 39 modules with a median gene number of 274. Remarkably, 17 of the 30 MaTARs are residing within the same module. This particular module (“blue,” Table S4) is enriched for genes that are highly expressed in our RNA-seq datasets, both in the normal mammary gland organoids as well as mammary tumors and hence was termed the “mammary epithelium” module. It is rather large, comprising 5,744 genes in total. Interestingly, only 2,588 of the module genes are classified as “protein coding,” emphasizing the tissue specificity of non-coding RNA species as well as their potential functional role.

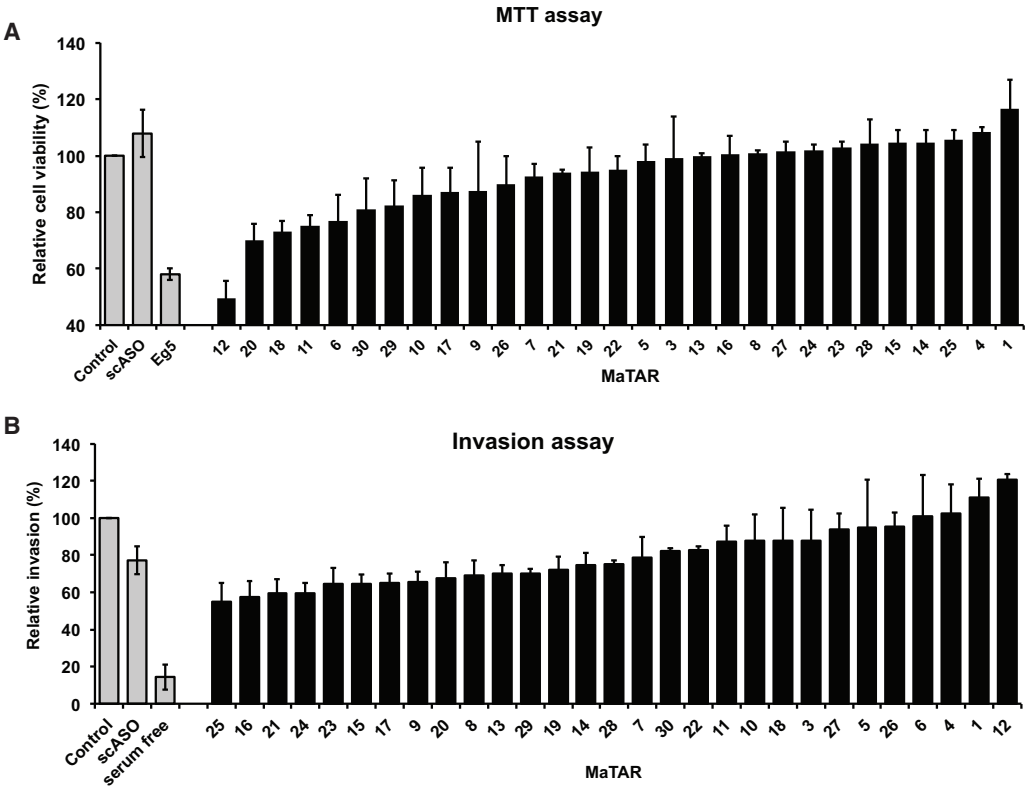
In order to identify key drivers of the module, we focused on potential “hub genes” with high intra-module connectivity (>1,550) and very good correlation with the module eigengene (≥0.9), marked in red in Figure 2H. At least six MaTARs (MaTAR5, 6, 9, 15, 16, and 30) fall into this region, indicating their importance within the module. Notably, 12 of the 17 MaTARs in the “mammary epithelium” module were also expressed higher in tumor tissue compared to normal organs (Figure 2G). This indicates that MaTARs are excellent marker genes and sufficient to stratify mammary epithelia from other tissue types. GO analysis of the module revealed association with several interesting biological processes that are in agreement with the top “hub genes,” including 33 ribosomal and small nucleolar RNA (snoRNA) genes, several components of the spliceosome, cyclooxygenase 2 and 3, but also typical marker genes for mammary epithelia like caseins and lactalbumin.

In summary, our findings demonstrate that we identified a module related to mammary epithelial development and/or mammary carcinogenesis and suggest that MaTARs are likely to play key roles in these functional processes.

Knockdown of MaTARs Decreases Cell Viability and Invasion

The results of our computational approach suggested that multiple MaTARs may contribute to mammary cancer progression. We tested this possibility by performing knockdown experiments of all 30 MaTARs using antisense oligonucleotides (ASOs) initially in primary MMTV-PyMT tumor cells. These oligonucleotides are short (20-mer), single-stranded DNA molecules containing phosphorothioate-modified nucleotides as well as modifications of the 2′-ribose (5-10-5 2′-MOE gapmer) (for review, see Geary et al., 2015). Importantly, we found that the uptake of ASOs in primary mammary tumor cells is efficient without the use of transfection agents, a mechanism that has been studied in detail in hepatocytes (Koller et al., 2011). One of the candidates, MaTAR2, was not expressed at levels consistent enough for reliable quantification using qRT-PCR and hence was excluded from knockdown experiments. For each of the remaining 29 MaTARs, we tested ten different specific ASOs and achieved knockdown efficiencies ranging from 38% to 89% after 24 hr using 5 μM of the most potent ASO (Figure S5A).

To initially investigate the functional impact of MaTAR downregulation on tumor cells, we combined the ASO treatment with cell viability or invasion assays (Figures 3A and 3B). Significant changes were defined as a consistent reduction ≥25% of viability or invasion compared to untreated control cells. Interestingly, MTT assays revealed that the knockdown of MaTAR12 leads to a 50% decrease in cell viability (Figure 3A), an effect comparable to ASO-mediated knockdown of Eg5 that served as a positive control (Koller et al., 2006). Knockdown of three additional candidates (MaTAR11, 18, and 20) resulted in ≥25% decrease of cell viability. We did not observe an effect on cell viability with a scrambled ASO control (scASO)



C

50% 100% 150%

	Viability	Invasion	Branching
MaTAR1	117	111	99
MaTAR2	n.a.	n.a.	n.a.
MaTAR3	99	88	83
MaTAR4	108	102	88
MaTAR5	98	95	83
MaTAR6	77	101	75
MaTAR7	93	79	87
MaTAR8	101	69	87
MaTAR9	88	65	82
MaTAR10	86	88	78
MaTAR11	75	87	72
MaTAR12	49	121	50
MaTAR13	100	70	80
MaTAR14	105	74	81
MaTAR15	105	65	85
MaTAR16	101	58	62
MaTAR17	87	65	61
MaTAR18	73	88	60
MaTAR19	95	72	71
MaTAR20	70	68	68
MaTAR21	94	59	86
MaTAR22	95	83	83
MaTAR23	103	65	84
MaTAR24	102	60	86
MaTAR25	106	55	56
MaTAR26	90	95	65
MaTAR27	102	94	77
MaTAR28	104	75	57
MaTAR29	83	70	78
MaTAR30	81	82	85

(legend on next page)

(Figure 3A). To confirm that our results obtained with the MTT assays are in fact detecting differences in cell viability, we performed assays measuring protease activity within living cells on a subset of the MaTARs and obtained similar results (Figure S5B).

In addition, we performed invasion assays of tumor cells in 96-well Boyden chamber plates upon ASO-mediated knockdown. We used serum-free medium as a positive control and detected a 25%–45% reduction of tumor cell invasion upon independent knockdown of 15 candidates (MaTAR8, 9, 13, 14, 15, 16, 17, 19, 20, 21, 23, 24, 25, 28, and 29; Figure 3B). While knockdown of several MaTARs impacts either the viability or the invasive potential of mammary tumor cells, indicating specific roles of the different non-coding transcripts in cellular processes, only MaTAR20 showed effects in both assays (Figure 3C). Of the 29 tested MaTARs, 11 lncRNAs (MaTAR1, 3, 4, 5, 6, 7, 10, 22, 26, 27, and 30) did not exhibit significant effects on cell growth or invasion upon knockdown.

As an example validation of the antisense knockdown experiments, we used the CRISPR/Cas9 system (Ran et al., 2013) to generate a knockout clone of MaTAR25, the lncRNA that exhibited the most dramatic impact on cell invasion (−45%, Figure 3B). Indeed, the genetic loss of MaTAR25 significantly reduced tumor cell invasion as well (−55%, Figures S5C–S5E), recapitulating the ASO-mediated knockdown. Our results indicate that our initial computational selection of upregulated lncRNAs identified a number of transcripts that have the potential to impact tumor cell growth or invasion.

Knockdown of MaTARs Inhibits Collective Cell Migration in Organoids

To further elucidate the functional role of MaTARs in a more physiological context, we performed ASO-mediated knockdown experiments in mammary tumor organoids (Barcellos-Hoff et al., 1989; Ewald et al., 2008; Fata et al., 2007; Nguyen-Ngoc et al., 2012). As for primary tumor cells, we found that mammary organoids perform transfection-free uptake of ASOs.

For the 29 tested MaTARs, we observed knockdown efficiencies in organoids ranging from 30%–68% after 6 days of treatment using 4 μM of the most potent ASO (Figure S6A). Untreated as well as scASO-treated MMTV-PyMT organoids generally exhibit branching in about 70%–75% of all organoids. Upon ASO-mediated knockdown, we detected a significant decrease (25%–50% less compared to untreated organoids) in branching morphogenesis for 11 candidates (MaTAR6, 11, 12, 16, 17, 18, 19, 20, 25, 26, and 28; Figure S6B). Interestingly, many of the MaTARs whose knockdown interfered with cell viability and/or invasion in 2D assays also exhibited a branching defect in organoids (Figure 3C). However, this correlation is not true for all MaTARs, indicating that 2D culture models different aspects of tumor cell growth than organoids.

As expected, we do not see phenotypic changes upon ASO-mediated downregulation for all MaTARs. While we obtained a knockdown efficiency of 55% for MaTAR8 (*ENSMUSG00000092569*, *Gm20544*, *lncRNA-Smad7*), we did not observe a loss of branching or any other phenotypic change in organoids (Figure 4D, additional images in Figure S6D). This is in agreement with previous findings describing that MaTAR8 is not involved in epithelial mesenchymal transition (EMT) (Arase et al., 2014). Our results indicate that different MaTARs might impact different aspects of tumor biology.

We further focused on MaTAR16 (*ENSMUSG00000086249*, *Gm12724*), MaTAR18 (*ENSMUSG00000085873*, *Ttc39aos1*), and MaTAR26 (*ENSMUSG00000097378*, *B230208H11Rik*), as these three transcripts showed very distinct organoid phenotypes upon ASO-mediated knockdown (Figure 4A). The three RNAs represent different transcript biotypes with MaTAR16 being classified as "processed transcript," MaTAR18 as "antisense," and MaTAR26 as "lincRNA." Both MaTAR16 and MaTAR26 are part of the "mammary epithelium" module, while MaTAR18 resides in another highly interesting module that is comprised of the gene-expression signature for embryonic stem cells and embryonic liver. This module is associated with GO terms related to replication and the cell cycle.

Knockdown experiments were performed with two different ASOs for each transcript to ensure that the observed phenotypes are not caused by off-target effects. Phenotypes for both ASOs are shown in comparison to organoids that were either untreated or treated with scASO (Figure 4A with additional images in Figure S6C). While knockdown of MaTAR16 and MaTAR26 completely abolished organoid branching, downregulation of MaTAR18 resulted in organoids without defined branches but with a "rough" surface. In addition, we observed that organoids treated with ASOs targeting MaTAR16 tend to be smaller compared to untreated or scASO-treated organoids (Figure 4A). Individual knockdown efficiencies are displayed in Figure 4B and correlate well with the loss of branching (Figure 4C). Compared to untreated controls, we detected a reduction in branching of 38% for MaTAR16, 40% for MaTAR18, and 35% for MaTAR26 with the best ASO, respectively.

To further elucidate the role of MaTAR16, MaTAR18, and MaTAR26, we performed ASO-mediated knockdown in control organoids derived from WT nulliparous mammary glands. Knockdown efficiencies of the ASOs are comparable to the experiments in tumor organoids (Figure S7A). Importantly, we do not see a pronounced loss of branching in the normal mammary

Figure 3. ASO-Mediated Knockdown of MaTARs in Primary Tumor Cells

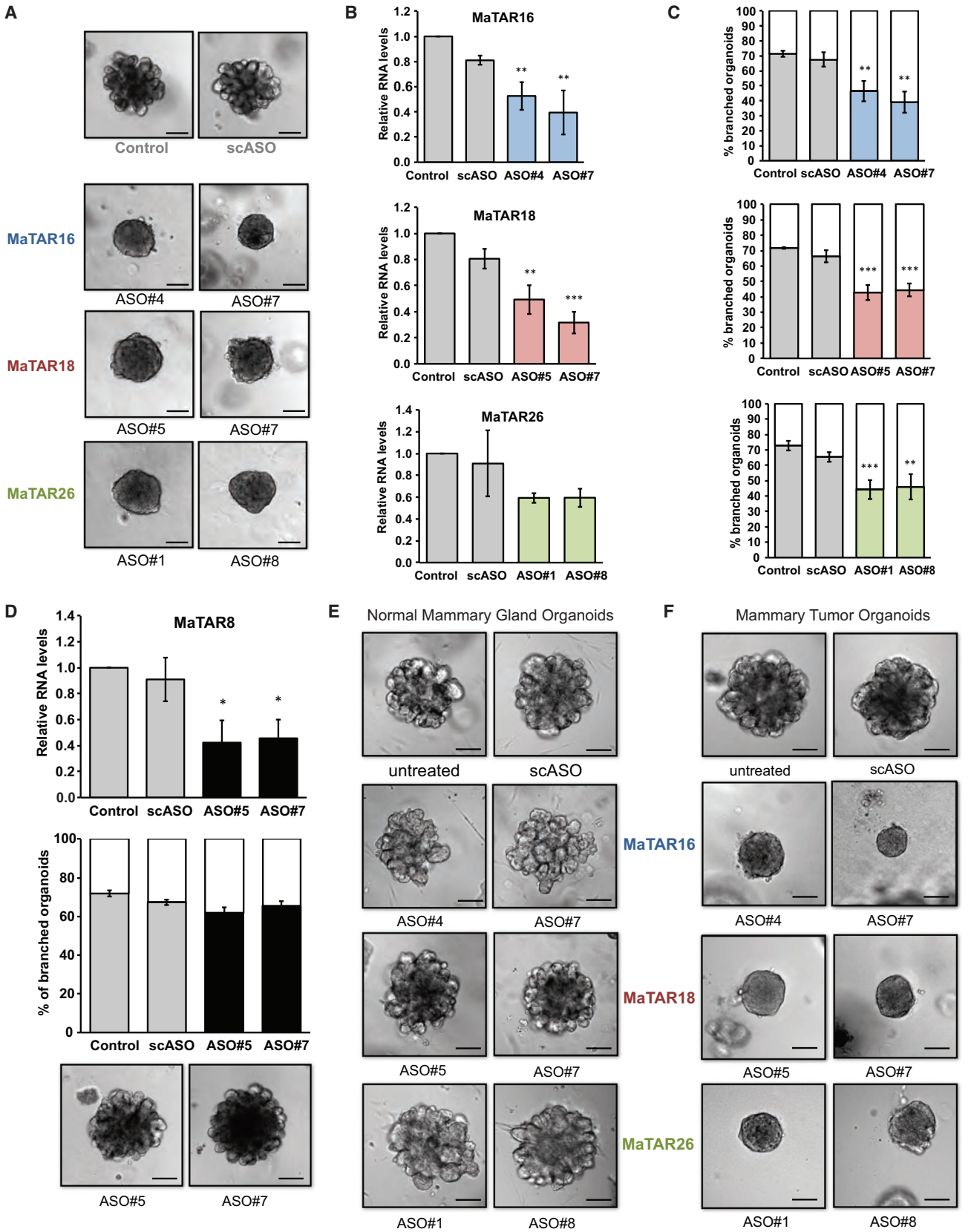
Primary mammary tumor cells were treated with 5 μM of the most potent ASO for 24 hr.

(A) MTT assay after ASO-mediated knockdown of MaTARs. Cell viability is normalized to untreated cells ("Control"). A scrambled ASO ("scASO") was used as a negative control; ASO knockdown of Eg5 was used as a positive control. Bars denote the mean of two replicates ± SE.

(B) Invasion assay after ASO-mediated knockdown of MaTARs. Like A, serum-free medium was used as positive control.

(C) Summary of observed effects on cell viability, cell invasion, and organoid branching upon ASO-mediated knockdown of MaTARs. Data are normalized to untreated cells (= 100%). Reduction of viability, invasive potential, or branching of ≥ 25% compared to untreated cells was defined as a significant difference. The color legend is indicated.

See also Figure S5.



(legend on next page)

gland organoids, strongly indicating the cancer-specific role of these three lncRNAs (Figure 4E). In addition, this experiment serves as another control, confirming that the observed reduction in branching morphogenesis of tumor organoids is not due to ASO induced toxicity or off-target effects.

To exclude the possibility that our results are dependent on the genomic background of the mice, we confirmed the findings for MaTAR16, MaTAR18, and MaTAR26 in organoids derived from C57Bl/6 MMTV-PyMT tumors (Figure 4F). Knockdown efficiencies and branching quantifications are shown in Figure S7B.

Human MaTARs Are Overexpressed in Breast Cancer

To confirm the relevance of MaTARs in breast cancer, we set out to identify human orthologs of all MaTAR genes. We compared mouse and human transcripts on the level of both sequence conservation and genomic location. If the mouse *MaTAR* sequence matched an annotated human gene using the BLAT alignment tool, then the respective transcript was defined as the human counterpart. For instance, BLAT results for *MaTAR1* indicate sequence identity in the human genome as well as a similar genomic context. It matches the last exon of a longer lncRNA (*LINC00461, ENSG00000245526*) in close spatial proximity of a miRNA gene (Figure 5A). The sequence of *MaTAR1* spans 2,419 bp in the human genome with a sequence identity of 92.6% and a BLAT score of 1438. This genomic region is highly conserved in vertebrates as indicated by the PhyloP conservation track (Figure 5A). The sequence-based approach is not informative for pseudogenes, as the parental protein-coding genes are the best sequence matches, e.g., *FAM96B* for a *Fam96b* pseudogene such as *hMaTAR2* (Table S5). As many non-coding RNAs are conserved between different species on the level of genomic location rather than based on sequence, possibly implying functional conservation (Diederichs, 2014; Ulitsky et al., 2011), we extended our analysis to synteny by analyzing the nearest neighboring genes of each *MaTAR* gene. The complete list of the genomic locations of potential *hMaTARs* is provided in Table S5. Interestingly, many *hMaTARs* are located in clusters of two or more non-coding genes in the genome, resulting in a total of 41 identified potential *hMaTARs*. Experimental validation will be necessary to identify the correct human counterpart unambiguously.

To evaluate the expression status of the potential *hMaTARs*, we analyzed RNA-seq data of The Cancer Genome Atlas (TCGA) by comparing breast tumors to matched normal tissues of 103 patients. Out of the 41 *hMaTARs*, we were able to obtain differential gene expression for 40 (Table S6). *hMaTAR3* dropped out due to insufficient read coverage. Notably, 28 *hMaTARs* were found to be upregulated in breast tumors, 19 of those with high statistical significance (adjusted $p < 0.05$), confirming that our initial RNA-seq screen in mouse models identified multiple lncRNAs that are potentially clinically relevant. We further investigated whether *hMaTARs* are expressed in a subtype-specific manner by analyzing all TCGA breast tumor samples with subtype information (463 datasets). Of the 19 upregulated *hMaTARs*, 11 were expressed significantly different ($q < 5.00E-11$, same range as PAM50 genes [Parker et al., 2009]) among the five TCGA subtypes "basal," "Her2-amplified," "luminal A," "luminal B," and "normal-like" (Table S7; Figure S8). In addition, *hMaTAR4* and *hMaTAR23* showed strong subtype specificity, indicating that these lncRNAs may be clinically relevant in certain subtypes despite not being significantly upregulated across the board (Table S6).

Here, we focus on *hMaTAR1*, as this lncRNA was upregulated the most in our original screen (80- and >100-fold in Her2-amplified and PyMT tumors, respectively). We found *hMaTAR1* to be significantly overexpressed in breast tumors compared to normal breast tissue (Table S6). We confirmed our analysis using previously published TCGA datasets (Yan et al., 2015) and detected a slight, but significant, upregulation in kidney carcinoma as well as substantial overexpression in lung squamous cell carcinoma (Figure 5B). Moreover, we observed statistically significant differences in the expression levels of *hMaTAR1* comparing different subtypes of breast cancer, with lowest levels in the luminal A and highest levels in the basal subtype (Figure 5C; Table S7). Notably, we also detected a strong correlation of high *hMaTAR1* levels and negative ER/PR status (Figure 5D; Table S7), matching the subtype data. Similar observations were made for seven additional *hMaTARs* (Table S7; Figure S9). Interestingly, the high expression levels in ER-/PR- tumors did not correlate with a genomic amplification of the *hMaTAR1* locus (data not shown), indicating that upregulation of this lncRNA may be caused by changes in gene expression and/or epigenetic factors. Survival analysis revealed that patients with low levels of

Figure 4. ASO-Mediated Knockdown of MaTARs in Mammary Organoids

Organoids were treated with 4 μM of specific ASOs for 6 days. Statistical significance was determined using a two-tailed, paired Student's t test; *$p < 0.05$, **$p < 0.01$, ***$p < 0.01$. Scale bar, 100 μm

(A) Differential interference contrast (DIC) images of tumor organoids upon knockdown with the two most potent ASOs targeting MaTAR16, MaTAR18, or MaTAR26 are shown. Knockdown of MaTAR16 and MaTAR26 leads to a loss of branching while knockdown of MaTAR18 results in organoids with tiny protrusions.

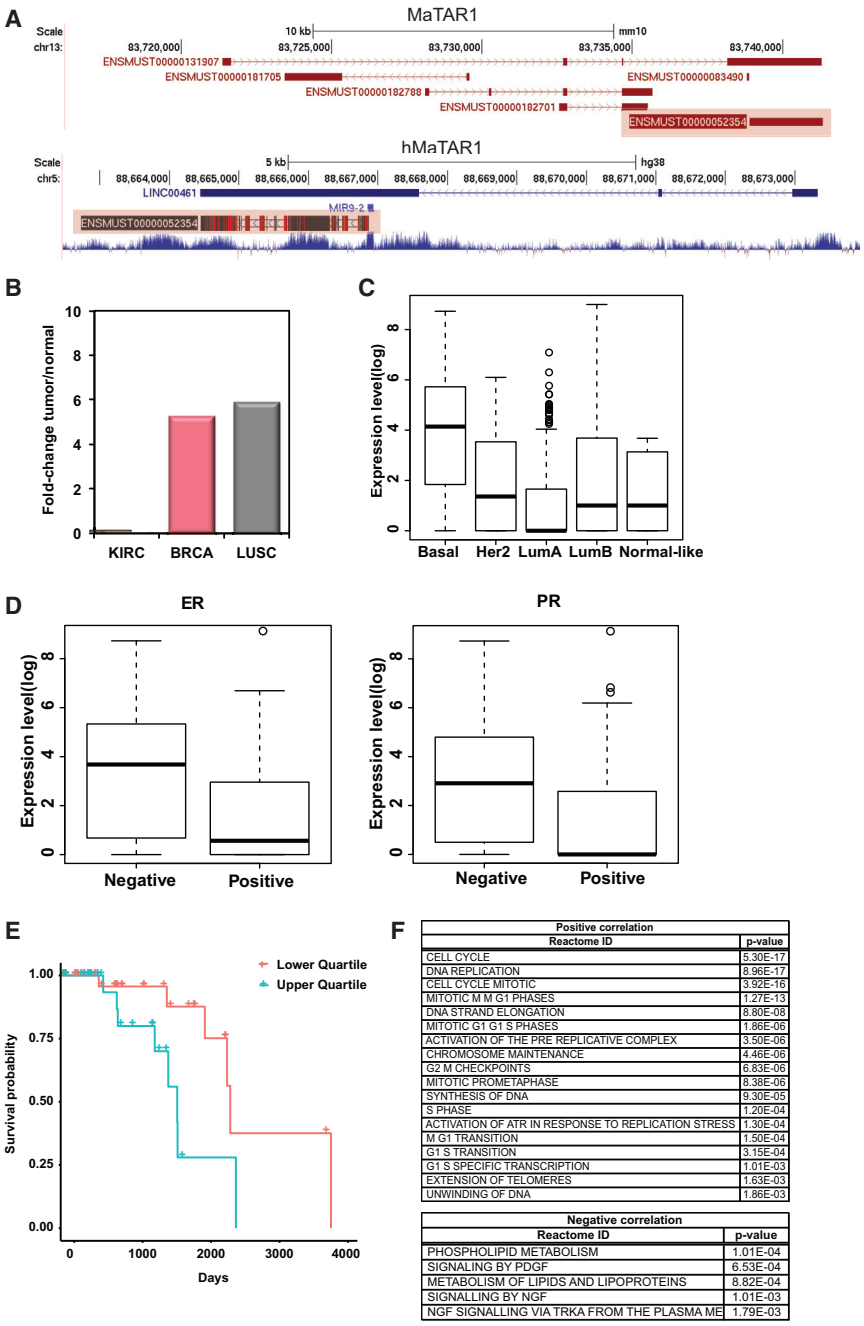
(B) qRT-PCR to determine the knockdown efficiency of two specific ASOs per MaTAR, normalized to untreated cells ("Control"). Bars denote the mean of three replicates ± SD.

(C) Quantification of tumor organoid branching upon knockdown with two specific ASOs per MaTAR. Reduction of organoid branching correlates directly with ASO-mediated knockdown efficiency. Black and gray bars indicate the percentage of branched organoids ± SD; white bars indicate the percentage of unbranched organoids. The mean of three biological replicates is shown; total number of assayed organoids per treatment = 300.

(D) Knockdown of MaTAR8 using two specific ASOs in tumor organoids. Upper panel: qRT-PCR to determine the knockdown efficiency. Bars denote the mean of three replicates ± SD. Middle panel: quantification of organoid branching. The mean of three biological replicates is shown; total number of assayed organoids per treatment = 300. Lower panel: DIC images of organoids upon knockdown.

(E) DIC images of WT normal mammary epithelia organoids upon knockdown with the two most potent ASOs targeting MaTAR16, MaTAR18, or MaTAR26 are shown.

(F) DIC images of C57Bl/6 MMTV-PyMT organoids upon knockdown with the two most potent ASOs targeting MaTAR16, MaTAR18, or MaTAR26 are shown. See also Figures S6 and S7.



Positive correlation	
Reactome ID	p-value
CELL CYCLE	5.30E-17
DNA REPLICATION	8.96E-17
CELL CYCLE MITOTIC	3.92E-16
MITOTIC M M G1 PHASES	1.27E-13
DNA STRAND ELONGATION	8.80E-08
MITOTIC G1 G1 S PHASES	1.86E-06
ACTIVATION OF THE PRE REPLICATIVE COMPLEX	3.50E-06
CHROMOSOME MAINTENANCE	4.46E-06
G2 M CHECKPOINTS	6.83E-06
MITOTIC PROMETAPHASE	8.38E-06
SYNTHESIS OF DNA	9.30E-05
S PHASE	1.20E-04
ACTIVATION OF ATR IN RESPONSE TO REPLICATION STRESS	1.30E-04
M G1 TRANSITION	1.50E-04
G1 S TRANSITION	3.15E-04
G1 S SPECIFIC TRANSCRIPTION	1.01E-03
EXTENSION OF TELOMERES	1.63E-03
UNWINDING OF DNA	1.86E-03

Negative correlation	
Reactome ID	p-value
PHOSPHOLIPID METABOLISM	1.01E-04
SIGNALING BY PDGF	6.53E-04
METABOLISM OF LIPIDS AND LIPOPROTEINS	8.82E-04
SIGNALLING BY NGF	1.01E-03
NGF SIGNALLING VIA TRKA FROM THE PLASMA ME	1.79E-03

Figure 5. hMaTARs Are Overexpressed in Breast Cancer

(A) Identification of a human MaTAR gene based on sequence identity. The upper panel represents the mouse genome (mm10); the lower panel represents the human genome (hg38). *MaTAR1* is highlighted (*ENSMUST00000052354*), ENSMUST00000083490 = miR-9-2.

(B) Expression fold change of *hMaTAR1* published TCGA RNA-seq datasets compared to the respective normal tissue. FDR < 0.00001. KIRC, kidney renal clear cell carcinoma; BRCA, breast invasive carcinoma; LUSC, lung squamous cell carcinoma.

(C) Differential gene expression of hMaTAR1 in subtypes of human breast cancer. Her2, HER2 amplified, LumA, luminal A; LumB, luminal B.

(D) Correlation of hMaTAR1 expression to estrogen receptor (ER, left) and progesterone receptor (PR, right) status.

(E) Survival curve for the luminal B subtype, comparing patients in the upper quartile of *hMaTAR1* expression to patients in the lower quartile of *hMaTAR1* expression.

(F) Reactome pathways positively or negatively correlated to *hMaTAR1* expression.

See also Figures S8 and S9 and Tables S5–S7.

Among these transcripts we prioritized 30 as MaTARs and showed that ASO-mediated knockdown of 20 MaTARs significantly reduces cell proliferation and/or invasion in 2D cell culture as well as branching in organoids. Phenotypic changes were only observed in organoids derived from mammary tumors but not normal mammary glands, indicating a specific role of MaTARs in cancer. A significant number of *hMaTARs* are overexpressed in human breast tumors and their expression signature correlates with subtype and hormone receptor status.

DISCUSSION

In recent years, high-throughput sequencing studies have begun to shed light on the emerging role of non-coding transcripts in cancer. Several studies suggest lncRNAs as novel prognostic biomarkers, tumor suppressors, and oncogenes (for review, see Cheetham et al., 2013; Huarte, 2015; Prensner and Chinnaiyan, 2011; Shore et al., 2012; Wahlestedt, 2013; Wapinski and Chang, 2011). However, many global gene expression studies were performed in human cancer cell lines. There are several potential caveats to using cell lines to study tumor biology, such as adaption to growth in vitro on plastic dishes, lack of corresponding normal tissue controls, cell immortalization, lack of a physiological ECM, and accumulation of chromosomal aberrations due to continued passaging. Importantly, transcriptome analyses revealed that commonly used

hMaTAR1 survive on average ~1,500 days (approximately 4 years) longer than patients with high levels of this lncRNA (Figure 5E). In addition, we performed pathway analysis of TCGA data and observed a significant positive correlation of pathways such as "cell cycle" and "DNA replication," hinting to a high proliferative rate of tumors with elevated *hMaTAR1* levels (Figure 5F).

In summary, we performed an RNA-seq screen to identify lncRNAs that are upregulated in mammary tumors and represent potentially oncogenic transcripts. We identified 109 lncRNAs that are upregulated at least 2-fold in MMTV-PyMT and 207 that are upregulated at least 2-fold in MMTV-Neu-NDL tumors.

cancer cell lines often do not represent tumors well enough and that tumors cluster closer to their corresponding normal tissues than to cell lines (Domcke et al., 2013; van Staveren et al., 2009). Primary 3D ex vivo organoids resemble the organization and physiology of native epithelia much better than cancer cell lines grown in 2D, and additionally model interactions with the ECM (Boj et al., 2015) (for review, see Clevers, 2016; Sachs and Clevers, 2014; Shamir and Ewald, 2014). To investigate whether organoids mirror the tumor on the level of global gene expression as well, we performed the first RNA-seq screen comparing the transcriptome of organoids to the mammary tumors from which they were derived. Importantly, our data demonstrate that non-coding RNA levels seem to remain virtually unchanged. Our findings support the potential use of patient-derived organoids as a model system to study tumor biology and to model drug responses ("personalized medicine" [Boj et al., 2015]; for review, see Lancaster and Knoblich, 2014; Shamir and Ewald, 2014).

We performed an unbiased RNA-seq screen to categorize the compendium of differentially regulated lncRNAs in mammary tumors compared to normal mammary epithelial cells. While we confirmed several known cancer-associated lncRNAs such as lincRNA-p21 in our screen, the majority of the 764 differentially expressed candidates have not been described previously in the context of cancer. Our study emphasizes the importance of unbiased screening approaches and represents a valuable resource of lncRNAs to pursue in regard to their involvement in mammary carcinogenesis and as prognostic and/or therapeutic targets.

Genome-wide gene expression studies often lack systematic functional validation. Here, we characterized the identified mammary tumor associated lncRNAs using both computational approaches and molecular assays. Motif analysis of all differentially regulated lncRNAs in our study revealed specific regulators for up- and downregulated non-coding RNA genes. Repressed lncRNAs show an enrichment in their promoter regions for the transcription factor Kaiso, which has been described to act as an oncogene by DNA methylation-dependent silencing of tumor suppressor genes in colon cancer (Lopes et al., 2008). Furthermore, we detect an enrichment of motifs for NF-κB and CEBP, two transcription factors that have been implicated in aggressive breast cancer and EMT (Bundy and Sealy, 2003; Huber et al., 2004). Upregulated lncRNA promoters are enriched for AR and p63 binding motifs. AR was described to stimulate breast tumor progression but its effect depends on the ER status of the patient (for review, see Chang et al., 2014). The transcription factor p63 is required to initiate collective cell migration and invasion in tumor organoids and in vivo (Cheung et al., 2013). Interestingly, our motif analysis revealed that the same regulators that control known tumor suppressor genes might downregulate the identified lncRNAs, while upregulated lncRNAs might be activated by factors that also control known oncogenes.

In addition to computational predictions, we performed ASO-mediated knockdown assays. Of the 290 upregulated candidates, we prioritized 30 lncRNAs, termed MaTARs, and observed that independent downregulation of 20 MaTARs leads to a significant reduction of cell proliferation, invasion, and/or collective cell migration. Given that the results of individual knockdown experiments were characteristic for the respective non-coding RNA, we propose that each MaTAR might impact different aspects of tumor cell growth and/or metastasis. Interestingly, knockdown of MaTAR16, MaTAR18, and MaTAR26 resulted in reduced branching of tumor organoids, but not normal mammary organoids. Our findings suggest that MaTARs might be dispensable for normal mammary epithelial growth but seem to be essential in the context of cancer.

Based on our extensive analysis in mouse, we propose that the human orthologs of MaTARs are likely impacting human breast cancer progression. In this regard, concerns are often raised about the lack of conservation between mouse and human lncRNAs (Derrien et al., 2012). Unlike protein-coding genes, the functions of lncRNAs are not necessarily directly linked to their nucleotide sequence but are often dependent on RNA secondary structure folding or syntenic transcription (Diederichs, 2014). Synteny between non-coding and protein-coding genes is frequently conserved across species (Ulitsky et al., 2011). Importantly, we identified potential human counterparts for all *MaTAR* genes and demonstrated that at least 50% of them are overexpressed in breast tumors compared to matched normal controls. The TCGA data analysis confirms that our RNA-seq screen in mouse models identified lncRNAs that act as important regulators of mammary tumor progression and whose functions may be conserved in the human system. Similar to the mouse studies, we revealed that *hMaTARs* can be expressed in a breast cancer subtype-specific manner or generally upregulated across all subtypes.

Of the 20 MaTARs that impact cell proliferation, invasion, and/or organoid branching, 11 have human orthologs that are upregulated in breast cancer. We suggest that these lncRNAs are likely drivers of tumor progression and could potentially be of clinical relevance. Future studies will reveal the detailed molecular mechanism by which *hMaTARs* act as well as their prognostic and therapeutic potential.

EXPERIMENTAL PROCEDURES

Organoid Culture

Organoids from wild-type nulliparous mammary glands, MMTV-PyMT tumors, and MMTV-Neu-NDL tumors were prepared and cultured as described previously (Ewald, 2013). For MMTV-PyMT mice, individual tumors were isolated and organoids were generated in a tumor-specific manner. Mammary epithelial fragments (organoids) were mixed with growth factor-reduced Matrigel at a concentration of 5 organoids/μL and plated as 80-μL domes in 24-well dishes. Organoids were grown in DMEM/F12 medium supplemented with 1× ITS (insulin, transferrin, and sodium selenite) media supplement, 1% penicillin/streptomycin, and murine FGF2 (2.5 nM). Animal experiments were carried out in the Cold Spring Harbor Laboratory Animal Shared Resource in accordance with Institutional Animal Care and Use Committee-approved procedures (see the Supplemental Experimental Procedures).

RNA Sequencing

Total RNA was isolated either directly from cryosections of the tumor tissue or from organotypic epithelial cultures using TRIzol according to the manufacturer's instructions. For tissue sections, the tumors were embedded in optimal cutting temperature (OCT) and cryosectioned. Sections from the middle of the tumor were stained using toluidine blue and assayed regarding the homogeneity of the section. Homogenous 30-μm sections comprising >90% malignant cells were immediately dispersed and homogenized in TRIzol. RNA quality was assayed by running an RNA 6000 Nano chip on a 2100 Bioanalyzer. For

high-throughput sequencing, RNA samples were required to have an RNA integrity number (RIN) $\geq$9. TruSeq (Illumina) libraries for poly(A)+ RNA-seq were prepared from 0.5–1 μg RNA per sample. To ensure efficient cluster generation, an additional gel purification step of the libraries was applied. The libraries were multiplexed (four to six libraries per lane) and sequenced paired-end 101 bp on the HiSeq2000 platform (Illumina), resulting in an average 40 Mio reads per library.

ASO-Mediated Knockdown in Primary Cells

MMTV-PyMT primary cells were seeded at a density of 20,000 cells/well (100 μL per well) into 96-well plates. Transfection-free uptake of ASOs was accomplished by adding 5 μM of either a MaTAR-specific ASO or scrambled ASO (scASO) to the primary cell culture medium immediately after seeding the cells. ASO sequences are provided in Table S8. Cells were incubated for 24 hr at 37°C and RNA was isolated using the RNeasy 96 kit (QIAGEN) according to the manufacturer's instructions. RNA samples were used directly in a one-step 384-well qRT-PCR (QuantiTect SYBR Green RT-PCR Kit, QIAGEN) on an Applied Biosystems 7900HT Fast Real-Time PCR System. qRT-PCR conditions were as follows: 30 min at 50°C for reverse transcription, 15 min at 95°C for the initial activation step followed by 40 cycles of 15 s at 94°C, 30 s at 60°C. *Peptidylprolyl isomerase B* (*cyclophilin B*) was used as an endogenous control to normalize each sample, and relative expression results were calculated using the $2^{-\Delta\Delta Ct}$ method. A list of primers used is provided in Table S9.

ASO-Mediated Knockdown in Organoids

Organoids were seeded at a density of 5 organoids/μL and plated as 80 μL domes in 24-well dishes. Transfection-free uptake of ASOs was accomplished by adding 4 μM of either a MaTAR-specific ASO or scASO to the organoid culture medium 15–20 min after the organoids were plated in Matrigel domes. Organoids were incubated for 6 days at 37°C and both medium, and ASOs were replenished at day 3. ASO sequences are provided in Table S8. Total RNA was isolated using TRIzol. DNase I treatment was performed for 15 min at room temperature (RT) to remove contaminating DNA. cDNA synthesis was carried out using TaqMan Reverse Transcription reagents (Life Technologies) and random hexamers according to the manufacturer's instructions. Real-time qPCR was performed using the Power SYBR Green Master Mix (Life Technologies) in 384-well plates using the Applied Biosystems 7900HT Fast Real-Time PCR System (Thermo Fisher Scientific). Cycling conditions were as follows: 10 min at 95°C followed by 40 cycles of 15 s at 95°C, 1 min at 60°C. Primer specificity was initially tested by agarose gel electrophoresis and subsequently monitored by melting curve analysis. *Peptidylprolyl isomerase B* (*cyclophilin B*) was used as an endogenous control to normalize each sample and relative expression results were calculated using the $2^{-\Delta\Delta Ct}$ method. A list of primers used is provided in Table S9. For visualization purposes and quantification of organoid branching, images were acquired using an Axio-Observer Live Cell inverted microscope (Zeiss) at 10× magnification.

TCGA Data Analysis

Analysis of TCGA data (Cancer Genome Atlas Network, 2012) was performed on the Cancer Genomics Cloud hosted by Seven Bridges Genomics (https://www.sbgenomics.com). For differential gene expression comparing 103 breast tumor samples to matched normal controls, raw RNA-seq reads were mapped to hg38 and counted as described above. The GENCODE v.24 gene transfer format (GTF) was used as a reference. Fold changes calculated for individual patients were averaged and statistical significance was calculated using Wilcoxon rank tests; p values were adjusted using the Benjamini and Hochberg method. Subtype-specific analysis, survival, and pathway analysis are described in detail in the Supplemental Experimental Procedures.

ACCESSION NUMBERS

The accession number for the RNA-seq data reported in this paper is NCBI GEO: GSE72823.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, nine figures, and nine tables and can be found with this article online at http://dx.doi.org/10.1016/j.celrep.2016.08.081.

AUTHOR CONTRIBUTIONS

S.D.D. and D.L.S. conceived the study, designed the experiments, and wrote the manuscript. S.D.D. performed the computational analyses of all mouse data. S.D.D. and K.-C.C. performed experiments and interpreted the data. S.D.D., J.S., O.E.D., and A.K. analyzed TCGA data. S.M.F., F.R., and C.F.B. designed and provided ASOs.

ACKNOWLEDGMENTS

We would like to thank all members of D.L.S.'s lab for helpful discussions and suggestions throughout the course of this study, especially Allen Yu for generously sharing his pSpCas9(BB)-2A-mCherry vector and Gayatri Arun for critically reviewing the manuscript. We acknowledge Mikala Egeblad, Senthil Muthuswamy (University of Toronto), and William J. Muller (McGill University) for providing mice and/or tumors. We acknowledge the CSHL Animal Shared Resource, Animal and Tissue Imaging Shared Resource, DNA Sequencing Shared Resource, and Microscopy Shared Resource (NCI 2P3OCA45508). This project was funded by NCI 5P01CA013106-Project 3 (D.L.S.), the Manhasset Women's Coalition Against Breast Cancer (S.D.D.), the Simons Foundation (J.S. and A.K.), NCI 5U01CA168409-03 (J.S and A.K.), and NCI 1U01CA188590-01A1 (A.K.). Access to TCGA controlled-access data were granted to A.K. by the Database of Genotypes and Phenotypes (dbGaP, project ID 1928). D.L.S. receives research support from and is a consultant to Ionis Pharmaceuticals.

Received: March 2, 2016
Revised: August 5, 2016
Accepted: August 23, 2016
Published: September 27, 2016

REFERENCES

Andrechek, E.R., Hardy, W.R., Siegel, P.M., Rudnicki, M.A., Cardiff, R.D., and Muller, W.J. (2000). Amplification of the neu/erbB-2 oncogene in a mouse model of mammary tumorigenesis. Proc. Natl. Acad. Sci. USA *97*, 3444–3449.

Arase, M., Horiguchi, K., Ehata, S., Morikawa, M., Tsutsumi, S., Aburatani, H., Miyazono, K., and Koinuma, D. (2014). Transforming growth factor-β-induced lncRNA-Smad7 inhibits apoptosis of mouse breast cancer JygMC(A) cells. Cancer Sci. *105*, 974–982.

Arun, G., Diermeier, S., Akerman, M., Chang, K.-C., Wilkinson, J.E., Hearn, S., Kim, Y., MacLeod, A.R., Krainer, A.R., Norton, L., et al. (2016). Differentiation of mammary tumors and reduction in metastasis upon Malat1 lncRNA loss. Genes Dev. *30*, 34–51.

Barcellos-Hoff, M.H., Aggeler, J., Ram, T.G., and Bissell, M.J. (1989). Functional differentiation and alveolar morphogenesis of primary mammary cultures on reconstituted basement membrane. Development *105*, 223–235.

Bergmann, J.H., and Spector, D.L. (2014). Long non-coding RNAs: modulators of nuclear structure and function. Curr. Opin. Cell Biol. *26*, 10–18.

Bergmann, J.H., Li, J., Eckersley-Maslin, M.A., Rigo, F., Freier, S.M., and Spector, D.L. (2015). Regulation of the ESC transcriptome by nuclear long noncoding RNAs. Genome Res. *25*, 1336–1346.

Bertone, P., Stolc, V., Royce, T.E., Rozowsky, J.S., Urban, A.E., Zhu, X., Rinn, J.L., Tongprasit, W., Samanta, M., Weissman, S., et al. (2004). Global identification of human transcribed sequences with genome tiling arrays. Science *306*, 2242–2246.

Boj, S.F., Hwang, C.-I., Baker, L.A., Chio, I.I.C., Engle, D.D., Corbo, V., Jager, M., Ponz-Sarvise, M., Tiriac, H., Spector, M.S., et al. (2015). Organoid models of human and mouse ductal pancreatic cancer. Cell *160*, 324–338.

Bundy, L.M., and Sealy, L. (2003). CCAAT/enhancer binding protein beta (C/EBPbeta)-2 transforms normal mammary epithelial cells and induces epithelial to mesenchymal transition in culture. Oncogene *22*, 869–883.

Cabili, M.N., Trapnell, C., Goff, L., Koziol, M., Tazon-Vega, B., Regev, A., and Rinn, J.L. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. *25*, 1915–1927.

Cancer Genome Atlas Network (2012). Comprehensive molecular portraits of human breast tumours. Nature *490*, 61–70.

Carninci, P., Kasukawa, T., Katayama, S., Gough, J., Frith, M.C., Maeda, N., Oyama, R., Ravasi, T., Lenhard, B., Wells, C., et al.; FANTOM Consortium; RIKEN Genome Exploration Research Group and Genome Science Group (Genome Network Project Core Group) (2005). The transcriptional landscape of the mammalian genome. Science *309*, 1559–1563.

Chang, C., Lee, S.O., Yeh, S., and Chang, T.M. (2014). Androgen receptor (AR) differential roles in hormone-related tumors including prostate, bladder, kidney, lung, breast and liver. Oncogene *33*, 3225–3234.

Cheetham, S.W., Gruhl, F., Mattick, J.S., and Dinger, M.E. (2013). Long noncoding RNAs and the genetics of cancer. Br. J. Cancer *108*, 2419–2425.

Cheung, K.J., Gabrielson, E., Werb, Z., and Ewald, A.J. (2013). Collective invasion in breast cancer requires a conserved basal epithelial program. Cell *155*, 1639–1651.

Chou, S.-D., Murshid, A., Eguchi, T., Gong, J., and Calderwood, S.K. (2015). HSF1 regulation of β-catenin in mammary cancer cells through control of HuR/elavL1 expression. Oncogene *34*, 2178–2188.

Clevers, H. (2016). Modeling Development and Disease with Organoids. Cell *165*, 1586–1597.

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. *22*, 1775–1789.

Diederichs, S. (2014). The four dimensions of noncoding RNA conservation. Trends Genet. *30*, 121–123.

Dimitrova, N., Zamudio, J.R., Jong, R.M., Soukup, D., Resnick, R., Sarma, K., Ward, A.J., Raj, A., Lee, J.T., Sharp, P.A., and Jacks, T. (2014). LincRNA-p21 activates p21 in cis to promote Polycomb target gene expression and to enforce the G1/S checkpoint. Mol. Cell *54*, 777–790.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. Nature *489*, 101–108.

Domcke, S., Sinha, R., Levine, D.A., Sander, C., and Schultz, N. (2013). Evaluating cell lines as tumour models by comparison of genomic profiles. Nat. Commun. *4*, 2126.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Ewald, A.J. (2013). Isolation of mouse mammary organoids for long-term time-lapse imaging. Cold Spring Harb. Protoc. *2013*, 130–133.

Ewald, A.J., Brenot, A., Duong, M., Chan, B.S., and Werb, Z. (2008). Collective epithelial migration and cell rearrangements drive mammary branching morphogenesis. Dev. Cell *14*, 570–581.

Fata, J.E., Mori, H., Ewald, A.J., Zhang, H., Yao, E., Werb, Z., and Bissell, M.J. (2007). The MAPK(ERK-1,2) pathway integrates distinct and antagonistic signals from TGFalpha and FGF7 in morphogenesis of mouse mammary epithelium. Dev. Biol. *306*, 193–207.

Geary, R.S., Norris, D., Yu, R., and Bennett, C.F. (2015). Pharmacokinetics, biodistribution and cell uptake of antisense oligonucleotides. Adv. Drug Deliv. Rev. *87*, 46–51.

Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M.-C., Hung, T., Argani, P., Rinn, J.L., et al. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature *464*, 1071–1076.

Guy, C.T., Cardiff, R.D., and Muller, W.J. (1992). Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease. Mol. Cell. Biol. *12*, 954–961.

Hansji, H., Leung, E.Y., Baguley, B.C., Finlay, G.J., and Askarian-Amiri, M.E. (2014). Keeping abreast with long non-coding RNAs in mammary gland development and breast cancer. Front. Genet. *5*, 379.

Huarte, M. (2015). The emerging role of lncRNAs in cancer. Nat. Med. *21*, 1253–1261.

Huarte, M., Guttman, M., Feldser, D., Garber, M., Koziol, M.J., Kenzelmann-Broz, D., Khalil, A.M., Zuk, O., Amit, I., Rabani, M., et al. (2010). A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell *142*, 409–419.

Huber, M.A., Azoitei, N., Baumann, B., Grünert, S., Sommer, A., Pehamberger, H., Kraut, N., Beug, H., and Wirth, T. (2004). NF-kappaB is essential for epithelial-mesenchymal transition and metastasis in a model of breast cancer progression. J. Clin. Invest. *114*, 569–581.

Iyer, M.K., Niknafs, Y.S., Malik, R., Singhal, U., Sahu, A., Hosono, Y., Barrette, T.R., Prensner, J.R., Evans, J.R., Zhao, S., et al. (2015). The landscape of long noncoding RNAs in the human transcriptome. Nat. Genet. *47*, 199–208.

Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Duttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermüller, J., Hofacker, I.L., et al. (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. Science *316*, 1484–1488.

Katayama, S., Tomaru, Y., Kasukawa, T., Waki, K., Nakanishi, M., Nakamura, M., Nishida, H., Yap, C.C., Suzuki, M., Kawai, J., et al.; RIKEN Genome Exploration Research Group; Genome Science Group (Genome Network Project Core Group); FANTOM Consortium (2005). Antisense transcription in the mammalian transcriptome. Science *309*, 1564–1566.

Koller, E., Propp, S., Zhang, H., Zhao, C., Xiao, X., Chang, M., Hirsch, S.A., Shepard, P.J., Koo, S., Murphy, C., et al. (2006). Use of a chemically modified antisense oligonucleotide library to identify and validate Eg5 (kinesin-like 1) as a target for antineoplastic drug development. Cancer Res. *66*, 2059–2066.

Koller, E., Vincent, T.M., Chappell, A., De, S., Manoharan, M., and Bennett, C.F. (2011). Mechanisms of single-stranded phosphorothioate modified antisense oligonucleotide accumulation in hepatocytes. Nucleic Acids Res. *39*, 4795–4807.

Kornienko, A.E., Guenzl, P.M., Barlow, D.P., and Pauler, F.M. (2013). Gene regulation by the act of long non-coding RNA transcription. BMC Biol. *11*, 59.

Lancaster, M.A., and Knoblich, J.A. (2014). Organogenesis in a dish: modeling development and disease using organoid technologies. Science *345*, 1247125.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics *9*, 559.

Lopes, E.C., Valls, E., Figueroa, M.E., Mazur, A., Meng, F.G., Chiosis, G., Laird, P.W., Schreiber-Agus, N., Greally, J.M., Prokhortchouk, E., and Melnick, A. (2008). Kaiso contributes to DNA methylation-dependent silencing of tumor suppressor genes in colon cancer cell lines. Cancer Res. *68*, 7258–7263.

Nguyen-Ngoc, K.-V., Cheung, K.J., Brenot, A., Shamir, E.R., Gray, R.S., Hines, W.C., Yaswen, P., Werb, Z., and Ewald, A.J. (2012). ECM microenvironment regulates collective migration and local dissemination in normal and malignant mammary epithelium. Proc. Natl. Acad. Sci. USA *109*, E2595–E2604.

Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., et al.; FANTOM Consortium; RIKEN Genome Exploration Research Group Phase I & II Team (2002). Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. Nature *420*, 563–573.

Parker, J.S., Mullins, M., Cheang, M.C.U., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., et al. (2009). Supervised risk predictor of breast cancer based on intrinsic subtypes. J. Clin. Oncol. *27*, 1160–1167.

Prensner, J.R., and Chinnaiyan, A.M. (2011). The emergence of lncRNAs in cancer biology. Cancer Discov. *1*, 391–407.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. *8*, 2281–2308.

Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. Annu. Rev. Biochem. *81*, 145–166.

Rinn, J.L., Kertesz, M., Wang, J.K., Squazzo, S.L., Xu, X., Brugmann, S.A., Goodnough, L.H., Helms, J.A., Farnham, P.J., Segal, E., and Chang, H.Y. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. Cell *129*, 1311–1323.

Sachs, N., and Clevers, H. (2014). Organoid cultures for the analysis of cancer phenotypes. Curr. Opin. Genet. Dev. *24*, 68–73.

Shamir, E.R., and Ewald, A.J. (2014). Three-dimensional organotypic culture: experimental models of mammalian biology and disease. Nat. Rev. Mol. Cell Biol. *15*, 647–664.

Shore, A.N., Herschkowitz, J.I., and Rosen, J.M. (2012). Noncoding RNAs involved in mammary gland development and tumorigenesis: there's a long way to go. J. Mammary Gland Biol. Neoplasia *17*, 43–58.

Siegel, P.M., Ryan, E.D., Cardiff, R.D., and Muller, W.J. (1999). Elevated expression of activated forms of Neu/ErbB-2 and ErbB-3 are involved in the induction of mammary tumors in transgenic mice: implications for human breast cancer. EMBO J. *18*, 2149–2164.

St Laurent, G., Shtokalo, D., Tackett, M.R., Yang, Z., Eremina, T., Wahlestedt, C., Urcuqui-Inchima, S., Seilheimer, B., McCaffrey, T.A., and Kapranov, P. (2012). Intronic RNAs constitute the major fraction of the non-coding RNA in mammalian cells. BMC Genomics *13*, 504.

Sun, M., Gadad, S.S., Kim, D.-S., and Kraus, W.L. (2015). Discovery, Annotation, and Functional Analysis of Long Noncoding RNAs Controlling Cell-Cycle Gene Expression and Proliferation in Breast Cancer Cells. Mol. Cell *59*, 698–711.

Tomita, S., Abdalla, M.O.A., Fujiwara, S., Matsumori, H., Maehara, K., Ohkawa, Y., Iwase, H., Saitoh, N., and Nakao, M. (2015). A cluster of noncoding RNAs activates the ESR1 locus during breast cancer adaptation. Nat. Commun. *6*, 6966.

Ulitsky, I., Shkumatava, A., Jan, C.H., Sive, H., and Bartel, D.P. (2011). Conserved function of lincRNAs in vertebrate embryonic development despite rapid sequence evolution. Cell *147*, 1537–1550.

van Staveren, W.C.G., Solís, D.Y.W., Hébrant, A., Detours, V., Dumont, J.E., and Maenhaut, C. (2009). Human cancer cell lines: Experimental models for cancer cells in situ? For cancer stem cells? Biochim. Biophys. Acta *1795*, 92–103.

Vikram, R., Ramachandran, R., and Abdul, K.S.M. (2014). Functional significance of long non-coding RNAs in breast cancer. Breast Cancer *21*, 515–521.

Wahlestedt, C. (2013). Targeting long non-coding RNA to therapeutically upregulate gene expression. Nat. Rev. Drug Discov. *12*, 433–446.

Wang, L., Park, H.J., Dasari, S., Wang, S., Kocher, J.P., and Li, W. (2013). CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. Nucleic Acids Res. *41*, e74.

Wapinski, O., and Chang, H.Y. (2011). Long noncoding RNAs and human disease. Trends Cell Biol. *21*, 354–361.

Xing, Z., Lin, A., Li, C., Liang, K., Wang, S., Liu, Y., Park, P.K., Qin, L., Wei, Y., Hawke, D.H., et al. (2014). lncRNA directs cooperative epigenetic regulation downstream of chemokine signals. Cell *159*, 1110–1125.

Yan, X., Hu, Z., Feng, Y., Hu, X., Yuan, J., Zhao, S.D., Zhang, Y., Yang, L., Shan, W., He, Q., et al. (2015). Comprehensive Genomic Characterization of Long Non-coding RNAs across Human Cancers. Cancer Cell *28*, 529–540.

Yang, F., Zhang, H., Mei, Y., and Wu, M. (2014). Reciprocal regulation of HIF-1α and lincRNA-p21 modulates the Warburg effect. Mol. Cell *53*, 88–100.

Zhang, X., Lian, Z., Padden, C., Gerstein, M.B., Rozowsky, J., Snyder, M., Gingeras, T.R., Kapranov, P., Weissman, S.M., and Newburger, P.E. (2009). A myelopoiesis-associated regulatory intergenic noncoding RNA transcript within the human HOXA cluster. Blood *113*, 2526–2534.

Zhang, X., Weissman, S.M., and Newburger, P.E. (2014). Long intergenic non-coding RNA HOTAIRM1 regulates cell cycle progression during myeloid maturation in NB4 human promyelocytic leukemia cells. RNA Biol. *11*, 777–787.